

# 2007 Census Bilingual Form Study

## Study Plan

Jennifer Tancreto and Joan Hill  
Decennial Statistical Studies Division

### 1. Introduction

The Census Bureau tested a bilingual census form in the 2005 National Census Test (NCT). The bilingual questionnaire had a “swim lane” design that provided two response columns, one in English and one in Spanish, each containing the same questions and response categories. This form was mailed to 10,000 housing units across the United States.

Results from the 2005 NCT show that the bilingual form significantly increased the self-response rate nationally (by 2.2 percentage points for paper response, and 1.1 percentage points for total response), and more specifically, in areas where there is a high concentration of non-White or Hispanic populations (Bouffard and Tancreto, 2006). Moreover, the bilingual form resulted in a higher proportion of Hispanic persons than the English-only form. However, item nonresponse rates for the bilingual form were higher for all household-level items and Hispanic origin compared to the English-only form (Bouffard *et al.*, 2006). There are many potential reasons for these item nonresponse discrepancies, including forms design, question wording, translation, and differences in the responding population.

The purpose of a follow-up study of the bilingual form is to determine if the item nonresponse issues can be resolved by improved form design and utilization of the questions that will be on the form in the 2008 Dress Rehearsal. Moreover, this test provides the opportunity to study the impact of the bilingual form in areas that contain a heavy concentration of Spanish-speaking people with limited English proficiency.<sup>1</sup> Furthermore, this study provides the opportunity to verify the 2005 NCT finding of increased response to the bilingual census form without the limitations associated with the 2005 NCT.<sup>2</sup>

### 2. Assumptions

#### 2.1 Universe

This test is intended to study the impact on unit and item nonresponse of mailing bilingual census forms to housing units in areas where many Spanish-speaking people with limited English

---

<sup>1</sup> People with “limited English proficiency” are those that report that they speak English “not well” or “not at all.”

<sup>2</sup> The 2005 NCT was a national test, while this follow-up study will target areas where the dual language design is likely most needed. While the results from 2005 cannot be directly generalized to this universe, we have no reason to suspect that the bilingual form would not maintain higher response than the English-only form among the population we are targeting.

proficiency reside. The forms distributed in this study (including an English-only control form) will be sent to these targeted areas.

## 2.2 Response Mode

We are conducting this study in the paper self-response mode only, since we are seeking information on bilingual design alternatives specific to that mode.

## 2.3 Uncontrolled Variation

The bilingual design has certain inherent characteristics that trigger variation from an English-only form. For instance, the bilingual form for the 2005 NCT was a booklet that was not folded in half like the English-only form. Therefore, it was mailed in a larger envelope than the English form. Given these types of variations, we will be unable to determine what role, if any, these factors play in any significant differences in the evaluation measures between the bilingual forms and the English-only form. That is, we will evaluate the bilingual forms with all of these changes as packages, and will be unable to tease out the individual effects of any one factor.

# 3. Methodology

## 3.1 Panel design

This study will include two bilingual form panels, and one English control panel. Each panel will have the same questionnaire content (i.e., the content selected for dress rehearsal). The control panel will be the English-only dress rehearsal form. The two experimental panels will contain variations of the bilingual form design to try to improve on the design that was used in the 2005 NCT. The first experimental panel will use a bilingual form with a swim lane format similar to the 2005 test design but with updated content. This form will include the Director's letter as the first page of the questionnaire booklet, like the 2005 test. The second experimental panel will contain a similar swim lane bilingual form; however, the first page of this booklet will be the first page of the questionnaire (i.e., "start here" with household questions). To accommodate the two additional pages at the end of the booklet that result from removing the director's letter from the cover, we will collect full information for persons 7 and 8, with a continuation roster for persons 9 through 12.

Note that this panel design allows for a clean test of our hypothesis that the presence of the director's letter on the front of the bilingual form booklet may distract from the household-level questions that are on the next page, thus causing increased item nonresponse to these questions, as shown in the 2005 NCT bilingual form analysis.

### 3.2 Sample Design

Housing units in this study will be selected from the mailout universe.<sup>3</sup> We plan to target the mailing of the bilingual forms and the control panel to areas of the United States that contain heavy concentrations of Spanish-speaking people with limited English proficiency.<sup>4</sup> By targeting forms to these areas, we can cleanly compare our evaluation measures across the English-only and bilingual forms.

We expect to mail each form to roughly 10,000 housing units in these areas, for a total of 30,000 housing units in the study.

### 3.2 Mailings

Every housing unit in this test will be sent an advance letter as a first contact. The advance letter will inform households that they will soon receive a request to complete a questionnaire for the test. The advance letter for the bilingual panels will have the dual language design.

The second mailing will be the initial questionnaire package. Housing units will receive a paper questionnaire and a first-class postage-paid return envelope. Also included in the mailing package will be a letter from the Census Bureau Director that encourages households to respond. Households selected for the bilingual panels will receive one of the bilingual questionnaires, and households selected for the control panel will receive the English-only form.

The third mailing will be a reminder mailing. The reminder mailing will ask households to respond to the census study if they have not already done so. Note that the reminder mailing will be an English-only postcard for the English panel; however, we will use a letter for both bilingual panels since the addition of the Spanish translation requires more space than is available on the postcard.

The fourth and final mailing will be a targeted replacement questionnaire. A replacement questionnaire will be sent to all housing units that have not responded by a pre-determined date. Accompanying the questionnaire will be a letter from the Director urging response. Note that the replacement questionnaire will be the English form for all panels, including the bilingual panels. This mailing strategy is planned for dress rehearsal and the 2010 Census, since the replacement questionnaire workload and timing require an English-only replacement mailing.

### 3.4 Analytical Comparisons

We plan to compare the following measures to evaluate the effectiveness of the bilingual forms:

- Response rates (Section 3.4.1)

---

<sup>3</sup> While this test will be restricted to mailout areas for budgetary reasons, update/leave areas will be included in the bilingual form targeting in the Dress Rehearsal and the 2010 Census.

<sup>4</sup> The methodology and data source to be used in targeting these areas are under discussion.

- Item nonresponse rates (Section 3.4.2)
- Bilingual form response patterns (Section 3.4.3)
- Demographic characteristics (Section 3.4.4).

### 3.4.1 Response Rates

We will test for differences in the unit response rates among the panels. These comparisons will help determine if the bilingual forms increase response in the targeted areas relative to the English-only form. We will also study whether there are differences in unit response among the bilingual forms.

### 3.4.2 Item Nonresponse Rates

We will then study item nonresponse rates among the forms. We will calculate the item nonresponse rates for person-level data items (relationship, sex, age/year of birth, Hispanic origin, race, overcount) and household-level data items (household count, undercount, tenure, and phone number).<sup>5</sup> When computing rates for the bilingual forms, item nonresponse will take into account both language columns; that is, we will deem an item missing if there is no response in the English column and there is no response in the Spanish column.

Additionally, we will compare item nonresponse rates across the forms for Hispanic respondents to determine which form is less prone to item nonresponse for this population. Note that, in the absence of information regarding who completed the questionnaire in the household, we will use the characteristics of Person 1 to describe the respondent. Admittedly, past research suggests that Person 1 is not the respondent roughly 30 percent of the time (DeMaio *et al.*, 1990). Therefore, we will study item nonresponse for all persons by their Hispanic origin as well.<sup>6</sup> We expect these results to be similar to the results for Hispanic respondents, if the assumption of the respondent as Person 1 holds.

Next, we will compute item nonresponse rates for the Spanish and English language columns separately on the two bilingual forms. We will compare item nonresponse between the English and Spanish columns for each individual bilingual form. We will also compare item nonresponse rates for each language column between the bilingual forms. Note that these analyses require the assignment of persons or households to each language column to compute rates. That is, for households or persons where data are provided in the English and Spanish columns, we will assign the person or household to the language column that was used most heavily when we compute item nonresponse.<sup>7</sup> For example, if person 2 provided all data in Spanish with the exception of sex, which was provided in the English column, he or she will be counted among the people assigned to the Spanish column and will be included as item nonresponse for the sex question.

---

<sup>5</sup> Household data such as count, undercount, tenure, and phone number will be examined since the 2005 NCT showed significantly higher item nonresponse for the bilingual household items, as mentioned in Section 1. Based on the assumption that this was due to the design of the 2005 bilingual form, we hope to eliminate this item nonresponse difference with the alternative reformatted forms used in this study.

<sup>6</sup> The household item nonresponse rates will be calculated when Person 1 is Hispanic only, since we have no information on the Hispanic origin of the person who provided these items.

<sup>7</sup> If both columns were completed equally, we will randomly assign the person or household to one of the columns.

### 3.4.3 Bilingual Form Response Patterns

We will then study response patterns on the bilingual forms to get a better understanding of how people are using the forms. We will first compute the percentage of returns that used each language column, and compare these numbers across the bilingual forms.

In an attempt to determine if there was any confusion stemming from the design of the bilingual forms, we will compute the percent of returns that used both the English and Spanish languages in responding. We will compare these rates across the various bilingual forms. We will also study response patterns to determine how the two languages are being utilized on each form. If there are a sufficient number of returns that used both languages to respond, we will perform a more extensive analysis, such as looking at the demographic characteristics of these cases.<sup>8</sup>

### 3.4.4. Demographic Characteristics

We believe that the bilingual forms will be more appealing than the English-only form to certain people, especially those who are primarily Spanish-speaking. This component of the analysis will determine whether the persons listed on the bilingual returns differ from those listed on the English returns. Additionally, this analysis will determine if there are differences in characteristics among the various bilingual forms.

## 4. Data Requirements

Paper questionnaire returns may be keyed or electronically captured. This analysis requires the following data files (note that additional files are required for survey implementation):

- Sample control file – a file containing ID numbers for each sampled housing unit. Each record should indicate the experimental panel to which the housing unit was assigned and basic geographic information.
- Response file – a file containing the response data and ID number for each questionnaire return. For bilingual returns, this means capturing responses in both languages.
- Check-in file – a file containing check-in data for all returns.
- UAA check-in file – file containing check-in data for “undeliverable as addressed” (UAA) forms, for both the initial questionnaire and replacement questionnaire mailings.

This analysis will be conducted on raw data as there will be no imputation or editing performed on these data. The Decennial Statistical Studies Division (DSSD) will process the data to identify blank forms and select a primary return in the event that a household returns more than one complete form.

---

<sup>8</sup> Less than one percent of responding households used both language columns in the 2005 NCT (Bouffard *et al.*, 2006). We anticipate minimal cases with this pattern in this test as well.

## 5. Division Responsibilities

We plan to limit the outsourcing of tasks associated with this survey in order to save time/money, as well as to implement a contact strategy as close as possible to the dress rehearsal and 2010. Since it would not be feasible to have a contractor target the replacement questionnaire mailing, DSSD and the Decennial Management Division (DMD) will work with the National Processing Center (NPC) to perform the bulk of the tasks associated with this fourth contact in the mailing scheme. Although the scope of the contract work is yet to be determined, we will attempt to limit it to handling (possibly addressing) and sending mailing packages if internal staff is not able to provide those services.

Decennial Management Division will take the lead on providing general project management; drafting clearance submission to the Office of Management and Budget; providing contract management including drafting the statement of work; producing, monitoring, and providing status on the schedule; and handling budget aspects of the survey. They will also manage forms development and forms printing.

Statistical Research Division (SRD) will work with DMD and DSSD in designing the forms and performing cognitive testing.

The NPC will check-in the returns and UAA forms. They will also key or capture the response data, as well as perform pre-specified quality checks on those items. NPC may be involved in compiling, labeling, and sending all mailing packages/letters/postcards (this responsibility is currently under discussion).

Decennial Statistical Studies Division will draft the sampling specifications, select the sample, provide the address file for printing, conduct the analysis, document the results, and perform quality assurance steps on those deliverables. In addition, DSSD may be involved in selecting the targeted replacement sample and providing an address file for printing (still under discussion).

## 6. Milestone Schedule

<b>Task</b>	<b>Estimated completion date</b>
Baseline study plan	10/2006
Finalize questionnaires	4/2007
Census Day	7/1/2007
Response data file	9/2007
Present preliminary results	1/2007
Final report	3/2008

Note: The schedule dates presented in this plan are tentative.

## **7. Risks**

- The schedule for this study assumes that we will have finalized the mailout questionnaire content for the Dress Rehearsal by February 2007, and final forms for this test ready at the same time. If content decisions are delayed, the survey results and subsequent delivery of the final bilingual form content to the Decennial Response Integration System contractor will be delayed.
- The addition of other test objectives to this study could critically delay survey implementation or limit the utility of the findings by confounding the test design.

## **8. Issues that Need to be Resolved**

- What data source will be used to target the areas of interest in this study?

## **9. References**

Bouffard, J., and Tancreto, J. G. (2006) “2005 National Census Test: Bilingual Form Analysis,” 2010 Census Test Memoranda Series, No. 22, September 15.

DeMaio, Theresa J. and Bates, Nancy A. (1990) “Who Fills Out the Census Form?”, *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 584-589.